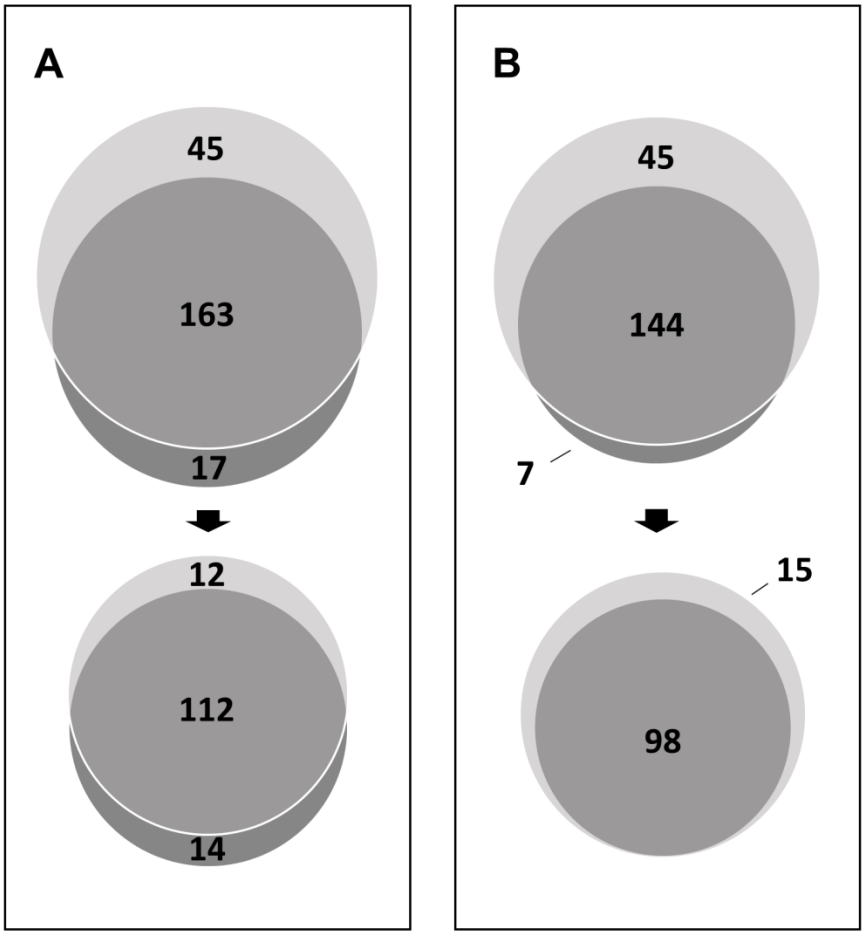
Supplementary for

High sensitivity crosslink detection coupled with integrative structure modeling in the Mass Spec Studio

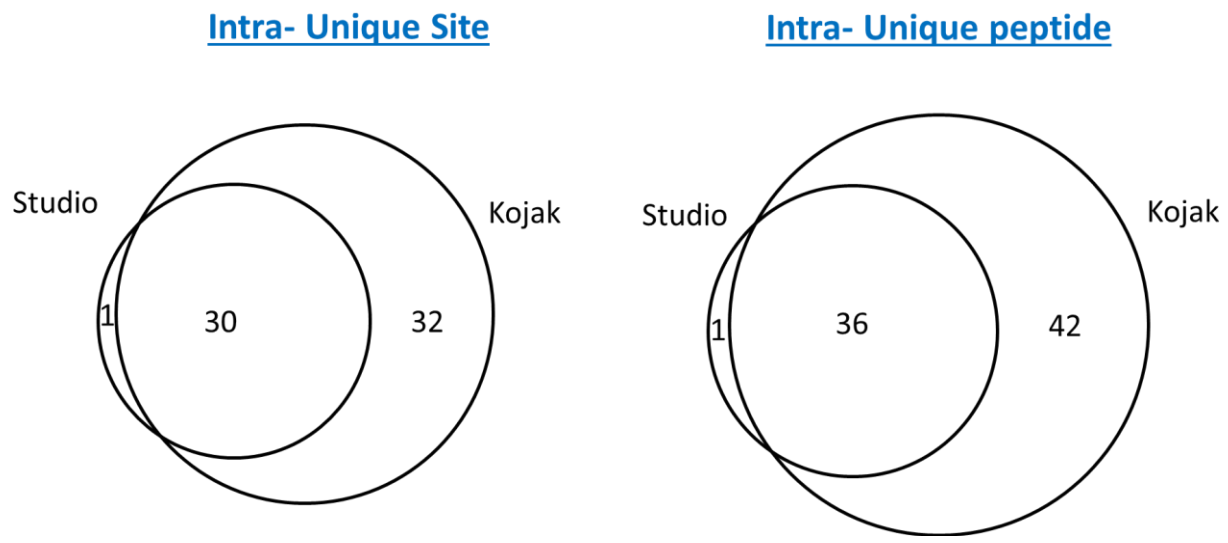
Vladimir Sarpe[1], *Atefeh Rafiei*[2], *Morgan Hepburn*[1] *Nicholas Ostan*[3], *Anthony B. Schryvers*[1,3] *and David C. Schriemer*[1, 2,*]

[1]Department of Biochemistry and Molecular Biology, [2]Department of Chemistry, [3]Department of Microbiology and Infectious Diseases, University of Calgary, Alberta, T2N 1N4, CANADA.

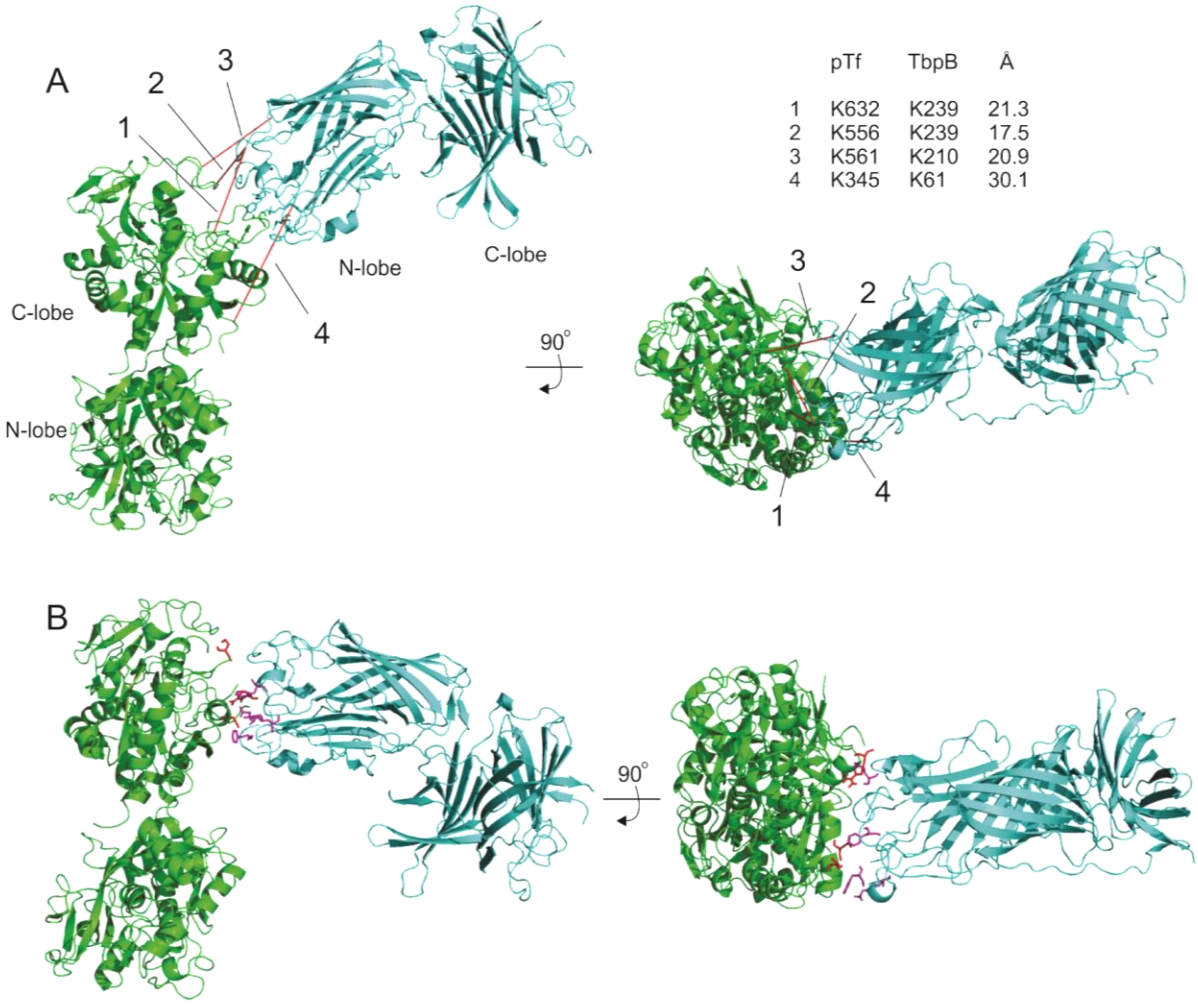
[*] correspondence to: dschriem@ucalgary.ca



Supplementary Figure 1. Comparing the mass-pairing and E-score methods of library reduction for crosslink detection sensitivity. (A) Venn diagrams showing overlap in hits using E-score (light gray) and mass-pairing (dark gray) reduction methods, based on a –ln transformed E′ score threshold of 13 applied in the final scoring. Top figure shows the full redundant counts, including multiple peptide charge states and bottom figure shows the unique subset of hits (B) Venn diagrams as in (A), with the application of an additional per-peptide –ln transformed $E_{\alpha/\beta}'$ score threshold of 10. Data were generated using a core library consisting of BSA + 10 *E. coli* proteins (1701 peptides) and a precursor tolerance window of 10 ppm. The E-score threshold for library reduction was 80% (1360).



Supplementary Figure 2. Comparing crosslink peptide identification in the Studio plug-in with Kojak, exclusively for the intraprotein crosslinks.
Venn diagrams showing overlap in identification using the Studio XL plug-in with library reduction by two peptide linear ion tags and Kojak with an open modification search. Both the uniquely determined linked sites are shown (left) and the unique peptides (right).Data were searched in the Studio using a 10/20 ppm tolerance window for both MS and MS^2 and recommended settings for Kojak (10 ppm for MS and 0.03Th for MS^2).



Supplementary Figure 3. Haddock-generated docking model using partial data sets.
(A) Side view and top view of a representative model of the top ranked cluster of solutions, generated when using only the four detected interprotein crosslinks found using the Studio plug-in. Orientation of TbpB and inaccurate alignment with mutation data do not support this model. Crosslinks shown as red bars. RMSD between right and left structures is 5.3Å (using pTf C-lobe and TbpB N-lobe) (B) Side and top view of a representative model of the top ranked cluster of solutions generated when using only the mutational data shown to abrogate binding of pTf to TbpB (see methods). RMSD between right and left structures is 4.9Å (using pTf C-lobe and TbpB N-lobe). TbpB is in cyan and pTf is in green, with the protein lobes indicated. RMSD calculated with ProFit.

Euclidean	Common	Kojak	Studio	**Surface**	Common	Kojak	Studio
<5	1	0	0	<5	0	0	0
5-10	4	5	0	5-10	4	0	0
10-15	11	12	0	10-15	8	8	0
15-20	5	7	1	15-20	4	9	0
20-25	2	1	0	20-25	6	5	0
25-30	2	1	0	25-30	1	1	0
30-35	1	3	0	30-35	1	2	0
40-45	0	0	0	40-45	0	2	0
45-50	0	0	0	45-50	1	0	0
50-55	0	0	0	50-55	1	2	1
55-60	0	0	0	55-60	0	0	0
60+	0	1	0	60+	0	1	0
Totals:							
<25	23	25	1	<25	22	22	0
>25	3	5	0	>25	4	8	1

Supplementary Table 1. Structural evaluation of the crosslink output. Distribution of Euclidean distances and surface-walk distances between linked residues, detected using either the Studio plug-in or Kojak, on the intraprotein crosslink data for pTf and TbpB. Calculations using www.xwalk.org, and distances in angstroms.